

ResViT: Residual Vision Transformers for Multimodal Medical Image Synthesis

Year: 2021 | Citations: 446 | Authors: Onat Dalmaz, Mahmut Yurt, Tolga Cukur

Abstract

Generative adversarial models with convolutional neural network (CNN) backbones have recently been established as state-of-the-art in numerous medical image synthesis tasks. However, CNNs are designed to perform local processing with compact filters, and this inductive bias compromises learning of contextual features. Here, we propose a novel generative adversarial approach for medical image synthesis, ResViT, that leverages the contextual sensitivity of vision transformers along with the precision of convolution operators and realism of adversarial learning. ResViT's generator employs a central bottleneck comprising novel aggregated residual transformer (ART) blocks that synergistically combine residual convolutional and transformer modules. Residual connections in ART blocks promote diversity in captured representations, while a channel compression module distills task-relevant information. A weight sharing strategy is introduced among ART blocks to mitigate computational burden. A unified implementation is introduced to avoid the need to rebuild separate synthesis models for varying source-target modality configurations. Comprehensive demonstrations are performed for synthesizing missing sequences in multi-contrast MRI, and CT images from MRI. Our results indicate superiority of ResViT against competing CNN- and transformer-based methods in terms of qualitative observations and quantitative metrics.